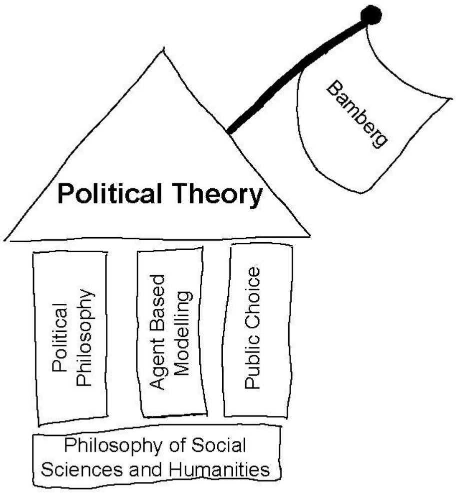
https://doi.org/10.20378/irb-108470

Networks of Polarisation

A Generative Mechanism

Daniel M. Mayerhoffer
Moritz A. Schulz
Simon Scheller
Jan Schulz

30 May 2025



Bamberg Contributions
to
Contemporary Political Theory

No. 4

Political Theory
Institute for Political Science
University of Bamberg
https://www.uni-bamberg.de/en/poltheorie/

Networks of Polarisation. A Generative Mechanism

D.M. Mayerhoffer[1,2], M.A. Schulz[3,4], S. Scheller[4], and J. Schulz[5,2]

[1]Amsterdam Institute for Social Science Research & Data Science Centre, University of Amsterdam
[2]Contribution co-funded by the German Federal Ministry of Education and Research (grant: DATI Pilot, e-lernen)
[3]Institute of Philosophy, FAU Erlangen-Nürnberg
[4]Institute for Political Science, University of Bamberg
[5]Economics Institute, University of Bamberg

Abstract

This paper presents a generative algorithm for simulating network polarisation based on attitudinal homophily, i.e., the tendency to connect to others with similar attitudes as oneself. To do so, it applies the notion of preferential attachment to node properties other than degree, aiding intuitive communication within and beyond the network science community. The algorithm works with one or more flexibly weighted attitude dimensions, heterogeneous populations. The generated networks commonly share features of real-world social networks such as (weak) small-worldiness. They can contribute to how-possibly explanations of stylised empirical facts, especially when empirical network data is missing. An example on polarisation in Germany on migration attitudes is used to illustrate this. We find that homophily is sufficient to create empirically observed polarisation as well as patterns of individual perception, such as viewing one's own opinion as moderate, over-estimating the actual level of societal, and dwindling open-mindedness towards others with different attitudes.

1 Introduction

Polarisation threatens the foundations of liberal democracy when political disagreement turns into mutual hostility. Social networks play a central role in this process; indeed, they are arguably what is getting polarised. Hence, network science provides tools to study how polarisation emerges and persists. Specifically, homophily — the tendency to link to others with similar views — is a likely candidate mechanism for generating polarised network structures [McPherson et al., 2001].

However, empirical study of these mechanisms is often hampered by a lack of empirical network data. To address this, we introduce a general-purpose generative algorithm that simulates attitudinal network formation through homophily. Its core idea resembles preferential attachment [Albert and Barabási, 2002], but applied to any node property, namely the similarity of nodes with regard to this property. The algorithm flexibly accommodates multiple attitude dimensions with different relative importance in tie formation, heterogeneous populations and directed/undirected links. It yields graphs that resemble empirically observed polarised networks in key properties. Hence, it provides a basis for how-possibly explanations of stylised empirical facts and is especially useful when empirical networks are unknown or partially observed.

The paper proceeds as follows. Section 2 clarifies key concepts surrounding polarisation. Section 3 introduces the model framework, including the homophilic linkage mechanism. Section 4 outlines how the model supports how-possibly explanations and can be validated against stylised facts, before exploring its use in simulating belief-updating. Section 5 presents an empirical case using migration attitudes in Germany. Section 6 concludes with a discussion of contributions and future directions both on the exemplary case and the algorithm in general.

2 Key Concepts

The concept of polarisation has no doubt made its way into the established vocabulary of describing and discussing politics, especially so in the US context: A Google search for the term "polarization" on the New York Times website returns 2,170 results for the year of 2024 alone (an average of six articles per day). As is perhaps unsurprising for a concept that has such currency both in public discourse and in several academic disciplines (political science, sociology, psychology, philosophy, media studies), it is not always clear whether the term is used consistently, or indeed what is denoted by it in the first place [Lelkes, 2016]. It is not, of course, the aim of this paper to contribute to explicating this concept and arbitrate among divergent uses. Still, it will be helpful to give some indication of the sort of phenomena which we understand may be studied using the proposed generative algorithm. To this end, we will in this section first suggest a core sense of polarisation operative across different contexts, and then briefly sketch what the term is usually further associated with when it is wielded as a timely diagnosis of putative problematic trends in Western democracies.

At its core, polarisation is a collective-level concept that picks out a typically bimodal distribution of relevant features within a population which combines *inter-group heterogeneity* with *intra-group homogeneity* [Mehlhaff, 2024]. So for this concept to be applicable, there must be at least one dimension on which individuals vary and individuals need to be attributable to groups (which may emerge from the distribution itself or may be identified independently, e.g. through party affiliation). When we say that the population is polarised with respect to this dimension, we imply two things: First, the groups significantly lie apart – for instance, Democrats have very different attitudes towards immigration than Republicans do. And second, members of either group are very much alike. This does not necessarily follow from the first: In theory, Republicans and Democrats could form non-overlapping sets with widely diverging means but still be spread out wide, such that there are significant disagreements among members of either group. Yet when we speak of polarisation, we typically have in mind that individuals gravitate towards tight-knit clusters on opposing ends of a spectrum.

In addition, it is usually taken for granted that the dimensions along which individuals vary are broadly speaking attitudinal or attitude-dependent features. For instance, we would not ordinarily hold that a given society is strongly polarised along the dimension of biological sex, even though it might be correct that biological males and females strongly resemble one another in sex-constitutive features (intra-group homogeneity) and both groups are in these respects quite unlike one another (inter-group heterogeneity). Rather, discussion of polarisation revolves around features such as normative beliefs or preferences (e.g., immigration policy, redistributive taxation, morality of abortion), empirical beliefs (e.g., dangerousness of vaccines, existence of climate change, theism), affective dispositions (e.g., hatred or fear of the outgroup, warm feelings towards the ingroup), and resultant behaviours (e.g., voting, social interaction).

Social psychologists have long studied subsets of these phenomena under the rubrics of *belief polarisation* and *affective polarisation.* In this literature, belief polarisation is understood to denote the phenomenon that deliberation in a like-minded group reliably causes the members' beliefs to become more extreme [Moscovici and Zavalloni, 1969, Myers and Bishop, 1970, Baron et al., 1996, Sunstein, 2002, 2009], where 'more extreme' is sometimes best construed as an increase in confidence in an existing belief and sometimes as the acquisition of a new belief that sits further along on the relevant attitudinal dimension, in the direction congruent with the group's established identity [Talisse, 2021, 213-15]. Information-based accounts of the underlying mechanism have highlighted how either kind of shift can be subjectively epistemically rational for group members given their exposure to corroborating evidence to the exclusion of countervailing arguments [Burnstein and Vinokur, 1977, Begby, 2022]. Such epistemic dynamics have garnered particular interest with regard to 'echo chambers' or 'filter bubbles' [Pariser, 2011, Steppat et al., 2022, Hobolt et al., 2024, Iyengar et al., 2018], especially on social media, which have in turn been conjectured to drive polarisation [Sunstein, 2018, Vaidhyanathan, 2018, Bruns, 2019, Nyhan et al., 2023]. Alternatively, the underlying mechanism has been theorised to accord with Social Identity Theory [Tajfel and Turner, 1986, Hornsey, 2008], whereby the mere fact of salient group membership is thought to propel an interest in ingroup cohesion and further distinction from the outgroup, such that members align their beliefs to those of others merely upon learning that the belief is characteristic of group members [Myers et al., 1980, Abrams et al., 1990, Lee, 2007, Anderson, 2021].

In turn, a concern with social identity is shared by scholars who study *affective polarisation*, typically associated with *partisanship* within the purview of political science [Huddy et al., 2015]: Whereas Social Identity Theory generally expects group membership to cause a favourable evaluation of the ingroup as compared to the outgroup [Billig and Tajfel, 1973], this literature is specifically interested in how political identity causes value-laden affective dispositions (such as fear or loathing of the outgroup) and discusses these under the rubric of polarisation [Iyengar and Westwood, 2015, Iyengar et al., 2019].[1]

This richer notion of polarisation as affect-laden and intertwined with partisan identities brings us closer to the term's mainstream use as a proliferating anxious diagnosis in political discourse [Iyengar, 2016, Finkel et al., 2020, Sides et al., 2018, Mason, 2018]. Here, it tends to be invoked when we observe a society-wide pattern of (increasing) inter-group heterogeneity and intra-group homogeneity across an (increasing) range of attitudinal and behavioural dimensions, which are newly associated with and more tightly integrated into a salient social – specifically, political – identity (causing *partisan sorting*). This effects that attitudes (or even aggregate identities) that were formerly intercombinable are now perceived to be incompatible, increasing the number of features people readily understand as signifiers of group membership. Accordingly, it has been associated with polarisation that attitudes on long-standing policy controversies become increasingly unidimensional [Hare, 2022]: One's views on immigration strongly predict their views on tax policy and on abortion rights. It is in this sense in which [Bakker and Lelkes, 2024, 419] hold that "polarisation ultimately refers to a collapse of dimensionality and the flattening of conflict". Yet beliefs beyond policy preferences equally seem to get implicated: For instance, your belief about whether a woman can have a penis might predict your belief about whether the sun was shining during President Trump's first inauguration, even though the two issues likely appear utterly unrelated to a naive outside observer [Schulz and Scheller, 2024]. In general terms, increasing pressures of intra-group homogeneity have plausibly been conjectured to co-opt contingent markers of group affiliation into constituting and maintaining the group identity [Mason and Wronski, 2018] – leading [Talisse, 2020, 84] to contend: "It's no stretch to say that today politics *simply is* our lifestyle." Such enriched and salient identities in turn impact preferences for social interaction beyond the realm of politics: Party affiliation has in the US become a major factor in the selection of potential romantic partners [Huber and Malhotra, 2017, Iyengar et al., 2018]; formerly apolitical fora of civil society such as churches or social clubs increasingly overtly identify as liberal or conservative [Talisse, 2020]; and certain professions skew along partisan lines [Bonica et al., 2016].

To summarise: When we talk about polarisation, we generally have in mind an attitude distribution that combines inter-group heterogeneity with intra-group homogeneity. When we worry about polarisation, we usually have in mind that polarisation clusters across dimensions into purified and affectively opposed identities.

3 Model Framework

3.1 Overview

The agent-based model developed in this study builds upon existing frameworks for network formation based on income homophily, as introduced in previous research [Schulz et al., 2022, Mayerhoffer and Schulz, 2022, Schulz and Mayerhoffer, 2023]. This model adapts the generative algorithm, allowing us to investigate polarisation and social division effects when network linkages are driven by homophily based on attitude(s). The network generation algorithm is versatile because it can be applied to any attributes or sets of attributes that can be quantified on a monotonous (preferably but not necessarily linear) scale. It is even possible to assess homophily based on a combination of attitude and demographics, such as gender or age. Furthermore, it can specify the number of links formed by a node (and hence its minimum degree), e.g. to capture different layers of social interaction [MacCarron et al., 2016].

---

[1] Indeed, Bakker and Lelkes [2024] admonish that the study of affective polarisation is squarely focussed on partisan identity and hardly at all on affect.

3.2 Agent Attitudes

The model simulates a population of N agents, each with a unique *attitude score* A_i for agent i, which is drawn from a continuous or discrete distribution representing the range of possible attitudes within a population. The attitude score A_i can reflect a single attitude dimension (e.g., political ideology) or can be extended to a vector of multiple attitude dimensions $\mathbf{A}_i = (A_{i,1}, A_{i,2}, \ldots, A_{i,m})$ if linkage based on their combination is of interest. For each agent, $\mathbf{A}_i$ should contain identical attitude dimensions, e.g., if migration attitudes are part of the model, all agents should be initialised with information on their migration attitude. Furthermore, attitudes need to be on (approximately) cardinal scale so that the homophilic linkage mechanism (cf. Section 3.3) can be applied to them. It is advisable to normalise all attitude dimensions, dividing all nodes' attitude values by the mean (of the respective dimension). This ensures that the segregation in the network depends only on the homophily (introduced below), and is not impacted by attitude scale.

Apart from these formal requirements, the notion of attitudes, as a technical one, may be operationalised in any way fit for the study in question. Attitudes can be taken from empirical data, other simulation models, or theoretically inferred distributions. Moreover, "attitude (dimension)" may refer to any node property, including demographics, on which one intends agents to form their links. (For the sake of simplicity, the remainder of this paper will refer to the relevant properties as attitudes.) If, for example, one would like to model homophily with regard to both age and beliefs about social security, $\mathbf{A}_i$ will consist of an age variable and one or multiple measures of the desired level of social security either as multiple variables (e.g., one on the level of unemployment benefits and another on social insurance contributions) or as a single measure combining them.

Any other basic agent characteristics concern their linking behaviour (see below). In addition to these basic properties, additional ones can be introduced, for example if nuanced interaction is to be modelled (cf. Section 4.5).

3.3 Link Formation Based on Attitudinal Homophily

The actual connection between agents follows a discrete choice mechanism, as outlined in Manski and McFadden [1981]. In a first step, an agent i calculates a weight $w_{i,j}$ for potential linkage with each other agent j. Based on these weights, the agent then draws d_i link-neighbours.

In the single-attitude case, the weight $w_{i,j}$ is given by:

$$w_{i,j} = \frac{1}{\exp(\rho_i \cdot |A_i - A_j|)} \tag{1}$$

where $\rho_i \in \mathbb{R}_0^+$ is the homophily parameter, determining the rate of decay in link probability with increasing attitudinal distance. Trivially, $w_{i,j} = w_{j,i}$ if $\rho_i = \rho_j$, and hence homogenous ρ implies fully symmetrical link weights.[2]

The key idea of the weighting function is to penalise differences in attitudes between i and j (shown in the denominator). The exponential nature of this penalty ensures that small differences receive only mild punishments, whereas large ones result in particularly low weights. As such, the weighting emphasises the homophilic tendency to pick others with close attitudes, making them much more likely as link-neighbours. Thereby, $\forall i \rho_i = 0$ implies equal weights for all agents, i.e., a simple random graph. For an increasing ρ, the weights of those others with great attitude differences shrink quickly (cf. [Schulz et al., 2022, Appendix A] for an analytical proof of this mechanism). Figure 1 showcases the effect of homophily in an example case.

In the multi-attitude case, the weight equation adapts as follows:

$$w_{i,j} = \frac{1}{\exp\left(\rho \cdot \sum_{k=1}^{m} \omega_{i,k} |A_{i,k} - A_{j,k}|\right)} \tag{2}$$

where the homophily parameter $\rho_i \in \mathbb{R}_0^+$ works as outlined above, and where $\omega_{i,k}$ is a weighting factor that determines the relative importance of attitude dimension k in agent i's overall homophily calculation. This weighted multi-dimensional approach allows to investigate the impact of different attitudes on network clustering and polarisation, adjusting the values of $\omega_{i,k}$ to emphasize or downplay specific dimensions. Like for the one-dimensional case, $\rho_i = \rho_j$ and $\forall k(\omega_{i,k} = \omega_{i,k})$ imply $w_{i,j} = w_{j,i}$.

Link weights are straightforwardly translated to probabilities when divided by the sum of weights that the agents assigns to all others. Based on these probabilities, i establishes their desired number of links d_i, using a multinomial sampling approach. If links are directed, it is more intuitive to think of agents forming their in-links (from the other agent to oneself) because the notion of the communication structure detailed below is one of individuals choosing whom to listen to rather than whom to address (consider, e.g., following others on social

---

[2] The algorithm is also straightforwardly applicable to j's attributes without direct comparison between i and j. This can be used to, for example, model people with extreme attitudes as especially (un)attractive link neighbours, or to cover linkage on demographic characteristics (such as a propensity to listen to those living in specific places).

media).[3] In that case, d_i specifies i's in-degree. For undirected links, in contrast, d_i specifies i's minimum degree for other agents can still select i as their link-neighbour if i has not selected them. Even after i has selected j as link-neighbour, i remains in the set of j's potential choices with no change to $w_{j,i}$. Should j also pick i, there still is simply one undirected link created[4] and neither of the agents picks an additional link-neighbour. Consequently, the order in which agents are picked for creating links is irrelevant to the resulting network (safe for the impact of randomness).

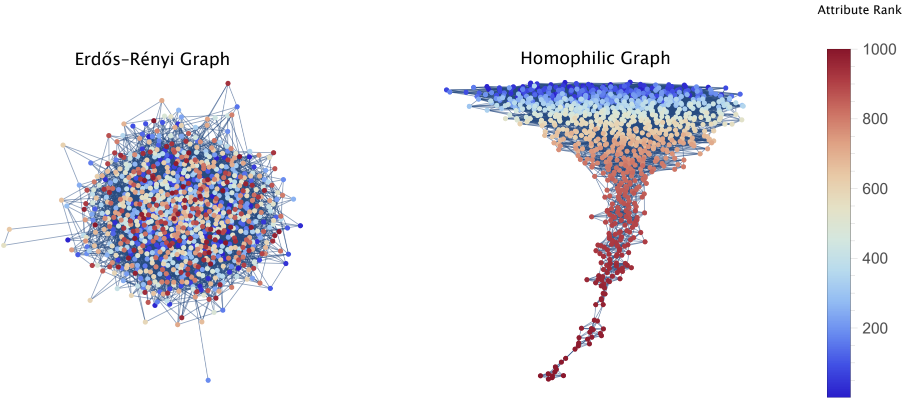


Figure 1: The effect of attribute homophily on network segregation. 1000 nodes, single log-normally distributed attribute with $\sigma = 1 \; \forall i d_i = 5 \wedge \rho_i = 10$, undirected graph. Visualisation uses Mathematica 13's "Spring Electrical Embedding" routine.

3.4 Communication and Sensing

With the network established, there are numerous possibilities to model agents' perceptions from or interactions on it. They range from simple sensing of immediate or higher degree neighbours' attitudes via updating of these attitudes or formation of attitudes on a new dimension to political action based on social influence. Section 4.5 details these options.

3.5 Accounting for (Random) Variation

Since attitudes, ω, ρ, and chance interact, drawing valid inferences (see Section 4.1 for details) from the networks generated by the algorithm requires preparation in setup. First, when working with theoretical input distributions, one might want to experiment with different distributional shapes. Second, exploring the empirically plausible combinations of attitude weights ω is necessary. Third, this also goes for values of ρ, where, as a rule of thumb, one might want to depart from a random network ($\rho = 0$), and increase ρ until the network reaches a level of segregation which is inadequate for the case studied (e.g., the network is no longer connected or the components get too small). Generally, the space to be explored depends on how much is known about the inputs. If, for example, attributes or their weights are known from empirical research (cf. Section 5 for an example), there is no need to vary these parameters, except for intentionally exploring counterfactuals.

Finally, to account for the impact of chance, one should have multiple networks simulated - do so-called Monte Carlo runs - for each combination of input values. Since the algorithm is computationally efficient, one might easily opt for a high number, such as at least 50, or even more if opting for heterogenous, randomly distributed values of ρ or ω. However, should one intend to run computationally demanding calculations on the resulting networks (such as calculating the clustering coefficient on larger graphs), working with fewer Monte Carlo runs might also be feasible, especially for homogeneous ρ and ω, where results tend to be fairly stable.

4 Application of the Network Simulation in Your Workflow

The generative algorithm proposed above alone does not answer any research questions, but the resulting networks can be used to do so in various ways, which are discussed in this section. Figure **??** provides an overview on how the different use cases relate to the model mechanism.

---

[3] Note, however, that formation of out-links is also possible and, of course, technically mostly equivalent.
[4] In a model variant, one could, of course, add higher weights to links that are established by both ends.

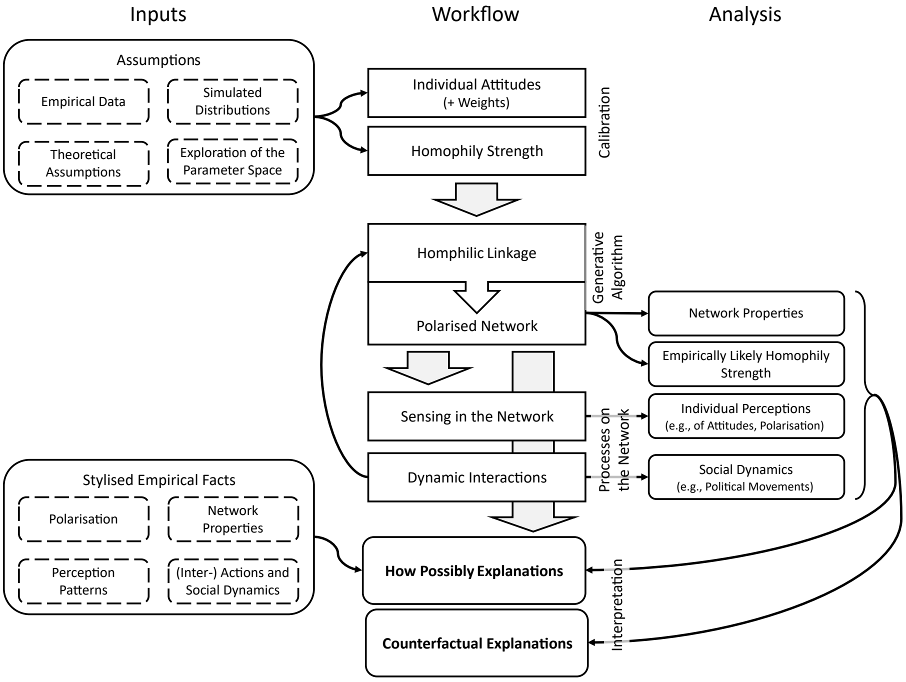


Figure 2: Workflow and use-cases of the homophily algorithm

4.1 How-Possibly Explanations of Stylised Empirical Facts

In the social sciences, stylised facts refer to simplified empirical regularities that capture broad, recurrent patterns across cases, abstracted from the noise and complexity of individual observations. Initially introduced by Kaldor [1961] in macroeconomics, the notion of stylised facts intends to allow theorists to construct models based on broad tendencies rather than the precise — and often imprecise — details of specific datasets. Kaldor saw stylising facts as a necessary abstraction to enable theory-building in the face of messy real-world data. Stylised facts typically represent unexplained empirical associations rather than fully elaborated causal claims [Hirschman, 2016]. For example, the persistent clustering of political attitudes within social networks, or the alignment of demographic characteristics and ideological positions, are stylised facts if they are consistently observed but not yet understood in terms of underlying mechanisms. In this sense, stylised facts are not endpoints of inquiry but starting points: empirical phenomena that call for theoretical explanation.

One response to stylised facts is through the use of how-possibly explanations. Unlike how-actually explanations, which aim to identify the true mechanism driving a phenomenon in a specific real-world case, how-possibly explanations offer candidate mechanisms that could plausibly account for the phenomenon. They help address the question: How could this empirical pattern conceivably come about? [Epstein, 1999, Reutlinger et al., 2018] In this context, simulation models are valuable tools for constructing how-possibly explanations. They develop ”parallel realities” [Sugden, 2009] where specific mechanisms operate in silico to reproduce stylised empirical patterns observed in real-world data. Such “surrogate systems” allow to study a target system where direct observation of the latter is not possible [?]. Our homophily-based generative algorithm serves this purpose. By simulating how agents form social connections based on attitudinal similarity, it reproduces network structures that mirror stylised facts of attitudinal clustering, segregation by political ideology, and associations between demographic and attitudinal characteristics. The match between simulated and observed patterns supports the plausibility of homophily as an explanatory mechanism — not as the definitive cause, but as a viable candidate within the broader landscape of social explanation.

4.2 Evaluation of the Generated Graphs

The networks generated by the algorithm can straightforwardly be evaluated through standard graph-theoretic metrics, i.e. clustering coefficient (to measure the degree of local cohesion within the network), degree distributions (to assess whether the network exhibits features such as right-skewness or hubs), modularity (to capture the extent of community formation and segmentation), and assortativity, particularly with respect to attitudes (to detect the emergence of polarisation), as well as connectivity and path length measures (to see how extreme segregation is). Of course, one can also use these basic properties to calculate compound ones, e.g., to assess small-world properties [Watts and Strogatz, 1998]. Sine the algorithm explicitly works on node attributes, the resulting graphs are attributed [Bothorel et al., 2015], and hence further case-specific methods should be considered.

These metrics may be complemented by visual inspections to gauge face validity. While a detailed calibration to empirical data is addressed in the following Subsection 4.3, in some cases a mere eyeballing might be enough to validate the graphs as internally coherent and externally in line with plausible social configurations.

Notably, the algorithm produces graphs that share key properties with Random Geometric Graphs, a well-studied class of spatial networks [Newman and Ziff, 2001, Dall and Christensen, 2002]. Like these, the networks generated by the present algorithm typically exhibit high clustering, right-skewed degree distributions, and a natural tendency towards local connection, all of which emerge from the underlying similarity-based linking rule. Firstly, this geometric structure aligns the generative model with observed topologies of real-world social and attitudinal networks, thereby reinforcing the plausibility of the simulated outcomes. Secondly, methods for studying Random Geometric Graphs are good candidates if one desires a deeper analysis of the graph structure.

If one does not know the features of an empirical network (e.g., due to lack of data), one can use the model to see which graphs are compatible with the input attitudes. Thereby, sensitivity analyses for a wide range of ρ and ω are necessary. The more information one has about the configuration of the corresponding empirical network, the easier it becomes to narrow down these values, as will be discussed next.

4.3 Gauging Empirically Likely Homophily Levels and Predicting the Effects of Changes in Homophily

Moving beyond face validity and sensitivity analysis, the model allows more rigorous validation by comparing the simulated networks' structural properties, attitude compositions, or behavioural patterns to empirical data or expectations derived from theory.[5] Networks generated under empirically plausible homophily parameters often exhibit weak small-world properties [Watts and Strogatz, 1998], high clustering, and distinct attitudinal clusters — features that are consistent with observed social and political network structures. Knowledge about how attitudes are distributed in empirical networks can help narrow down plausible values of ρ and ω.

This approach fits into what [Squazzoni, 2012, ch. 4.3] labels "multi-level empirical validation", which combines two logically distinct but substantively interdependent tasks: The first is empirical specification, the calibration of model inputs — here, the attitudinal distributions. The second is empirical validation of model outputs against stylised facts and known empirical patterns. Importantly, this process retains the explanatory orientation of the model. While not aiming to pinpoint the exact generative process in any specific empirical case (as in a how-actually explanation), the model's ability to replicate empirically grounded regularities supports its role as a how-possibly explanation. In this context, empirical homophily levels inferred from empirical data (e.g., attitudinal agreement among ego-alter pairs) can inform the model's parameter space. Conversely, the model can be used to estimate the homophily strength that would be required to generate a given observed degree of segregation or polarisation, offering a tool for counterfactual or policy-relevant scenarios.

Additionally, the model can be used to explore the effects of hypothetical changes in homophily — for instance, reductions brought about by deliberate mixing policies or increases due to rising affective polarisation. These experiments can reveal nonlinearities, tipping points, or emergent dynamics that would be difficult to anticipate through intuition alone.

4.4 Investigating the Complex Relationship Between Attitude Dimensions and Homophily

The model allows for the investigation of interdependencies between different attitude dimensions by simulating how homophily on one dimension may lead to segregation on another — even if the latter is not directly involved in link formation. One may simply achieve this by setting the homophily weight for the dependent dimension k to zero, i.e., $\forall i(\omega_{i,k} = 0)$. If the resulting network nonetheless exhibits significant clustering or assortativity on that dimension, this indicates that the segregation is an emergent byproduct of homophily on the other included dimensions. This approach might often be interesting to assess the impact of homophily in demographic features on attitudinal segregation, or vice versa.

Such analysis moves beyond detecting mere correlation between variables: it probes whether the network structure itself induces and amplifies these correlations. In particular, the model can reveal whether observed associations between attitudes reflect structural artefacts of network formation under homophily. This is especially pertinent, where the interplay between different homophily levels and attitude dimension weights becomes analytically intractable. The simulation framework offers a way to explore this interaction space systematically and to assess the extent to which network-driven co-segregation patterns may explain the empirical phenomena of affective polarisation, ideological bundling, or demographic echo chambers outlined in Section 2. With this, the model also presents a tool for highlighting the generative role of the network in shaping — rather than

---

[5] For the sake of simplicity and because we expect most of our readers to have this background, we will in the following only talk about empirical data. Yet, our explanations also cover an application of the simulation model to theory building [Taghikhah et al., 2021, Schwaninger and Grösser, 2008].

merely reflecting — observed attitudinal alignments, and for disentangling the structural from the substantive sources of such correlations.

4.5 Simulation of Processes on the Network

Simulated networks generated by the model can serve as surrogates for real-world social structures in cases where empirical network data are unavailable or incomplete [**?**]. In such scenarios, the simulated graphs provide a plausible basis for studying how beliefs or attitudes evolve through social influence processes embedded in homophilically structured environments. Moreover, these simulations can be used to explore counterfactuals or to test interventions aimed at reducing polarisation, such as changes in network openness or the introduction of bridging agents. The surrogate network approach thus offers a flexible framework for studying belief dynamics in a controlled yet socially plausible setting, grounded in empirically-informed network structures.

4.5.1 Basic Sensing as Social Sampling

The most basic set of processes to model on the network is social sampling [Galesic et al., 2012, Dasgupta et al., 2012], that is agents sensing others' attitudes. For example, they can situate their own attitude in their immediate social context of their ego network, compare their link-neighbours' attitudes to each other, or compute summary statistics of attitudes. Furthermore, simple inductive statistics, such as correlations between link neighbours' attitude dimensions, might be of interest. To assess the representativeness of the individual social samples, one can compare these local samples to the corresponding characteristics of the population as a whole.

Based on their sensing, agents can form beliefs about central tendencies in the population but also about how attitudes are distributed, and about properties of the network which they live on. For the purpose of investigating polarisation in networks, individual beliefs about this polarisation might be the prime characteristic; similarly, access to their link-neighbours' node properties would allow agents to gauge, for example, density or proportion of closed triads.

All the above can be taken only from the agents' own link-neighbours, but including neighbours of neighbours or mediate connections selected according to other criteria is equally possible. Likewise, agents can be made to retrieve characteristics other than attitudes, such as node properties or the aforementioned perceptions. The latter can be used to model meta-perceptions or iterative updating, as detailed in the following.

4.5.2 Dynamic Interaction

The iterative updating of beliefs about attitudes described above actually constitutes a first form of dynamic interaction: A simulation on the set network would run for multiple time steps and in each step, agents would formulate their new belief about attitudes, taking into account visible attitudes as well as a combination of one's own and others' beliefs from the previous step in a Bayesian updating process. Again, taking the ego-network, or including higher-degree neighbours, is the most straightforward way to define an agent's vision but other approaches are possible as well. For example, one could have each agent i select a random other j to retrieve information from, and each agent on the shortest path from j to i would also gain this information; such a mechanism would put more central nodes in an epistemically superior position.

One could also have agents utilise the information they receive over time to form more nuanced beliefs about the network topology and their own position within it. While the static version of this sensing described above would limit agents to extrapolation from their immediate context, learning about information passing through them or observing their neighbours' belief changes might also allow for insights into clusters that the assessing agent is not part of.

Probably the most obvious use-case for the attitudinal networks generated by the presented algorithm is to explore updating of the attitudes themselves, given initial polarisation. To that end, one could simply run Degroot [1974] on the network. Alternatively, one might use the network as a starting point for bounded confidence dynamics [Meylahn and Searle, 2024, Hegselmann et al., 2002, Deffuant et al., 2002]: Agents in the Deffuant-Weisbuch Model could select communication partners only from link-neighbours; those in the Hegselmann-Krause Model could have the size and shape of their confidence interval determined by the attitudes in their ego-network. By applying such updating mechanisms to the simulated networks, one can examine the consequences of different homophily strengths or network topologies for belief convergence, polarisation, or fragmentation. For instance, higher homophily may not only increase attitudinal clustering in the initial network but also hinder cross-cutting influence, thus stabilising or even intensifying initial divides.

Obviously, one can also have the network adapt over time to changes in attitudes to showcase feedbacks between opinion and network polarisation. Agents can drop links with a likelihood inferred from inverse weights and if they dropped a link form a new one according to the algorithm. This way, it is also possible to model polarisation from scratch, by having agents start from a random network. Likewise, one may want to study how attitudes and the network co-evolve over time, following e.g. [Henry et al., 2016].

Building on belief formation, one could also use the network to study individuals taking action in a polarised world. For example, to explain the fates of revolutions and street protests, the extant theoretical and simulation literature focuses mainly on individual decisions or attitude patterns [Kuran, 1989, Klein and Marx, 2017, Asgharpourmasouleh et al., 2019]. Running such simulations on homophilic networks can highlight the role of polarisation in this context: When agents see that their peers share their attitudes or deem change attainable they might feel motivated to take action, even if they are actually in the minority.

5 Exemplary Analysis: Migration Attitudes, Polarisation, and Perceptions in Germany

The purpose of this section is to illustrate potential use cases of the proposed generative algorithm. To do so, we generate polarised networks based on attitudes towards migration in Germany, and showcase how they can explain stylised empirical facts on actual and perceived polarisation in the German society.

5.1 Polarisation of Attitudes towards Migration in Germany

5.1.1 Relevant Stylised Empirical Facts

Attitudes toward migration have become highly polarised in many societies, which is to say opinions split into opposing camps with little middle ground [Alesina and Tabellini, 2024]. Polarisation on immigration is evident in anti-immigrant movements and contributes to the rise of far-right anti-immigration parties. Germany with its recently emboldened right-wing *AfD* party or the xenophobic *PEGIDA* movement, following the large influx of refugees in around 2015-16, is an example for this [Ueffing et al., 2015]. Windzio [2025] empirically identifies these two distinct clusters in German opinion data, and shows that similarity ties in polarising attitudes are strong within each cluster but weak between clusters. Hence, pro-immigration Germans tend to resemble each other (intra-group homogeneity) and differ starkly from those who are anti-immigration (inter-group heterogeneity), forming two coherent network components.

Communication plays a key role in polarisation. Coverage of the topic in traditional media [Schneider-Strawczynski and Valette, 2025] and social media interactions [Nasuto and Rowe, 2024, Esteve-Del-Valle, 2022] drive attitude development; due to its emotional dimension, the topic is likely to produce stronger narratives and information received will be more uniform on an intrapersonal level, yet less uniform on an interpresonal one; this, in turn, causes a more polarised attitudinal landscape. Polarisation is also salient for many people, with perceptions of polarisation are empirically upwards biased. Moreove, it is this perception rather than actual polarisation being related to (a lack of) trust and voting propensity [Enders and Armaly, 2019]. Furthermore, people believe their own attitude to be in line with the social norm [Dippel et al., 2022].

In sum, we derive three stylised facts relevant to our case:

1. People tend to perceive their own attitude as moderate, i.e., close to the mean and median of the attitude landscape.
2. People tend to overestimate the level of polarisation, i.e., they perceive their own social context to be more different from alters than is the case in comparison with society-wide estimates.
3. Segregation and polarisation cause lower levels of open-mindedness [Turner, 2023, Wollebæk et al., 2019]; hence, confidence intervals are small towards both individuals and political parties.

5.1.2 Data and Variable Construction

To model polarisation on migration attitudes representatively for Germany, this study draws on data from the German General Social Survey (ALLBUS) by GESIS-Leibniz-Institut Für Sozialwissenschaften [2024]. The 2021 wave, used in this analysis, employed a mixed-mode self-administered design (paper- or web-based) and covered a broad range of sociopolitical topics, including attitudes toward immigration and migration policy. Table 1 shows the variables used. Since questions measuring attitudes toward different migrant groups (mi05 to mi11) were only administered to respondents in a subsample, we only include this subsample (which, like the study as a whole is representative of the German populace) in our analysis.

We construct a composite migration scepticism index based on all migration-related variables in Table 1, i.e., all but *pa01.* This index is simply computed as the unweighted arithmetic mean of the recoded responses across all eleven items for each respondent (ignoring missing values). It is normalised, such that the population mean index is 1; values below 1 thus indicate above-average openness to migration, while values above 1 indicate below-average openness.

Figure 3 shows the distribution of the composite migration scepticism index, for each category on the self-description on a left-right scale (variable *pa01*). The latter will be used to create a benchmark for polarisation of

Variable	Question (English Translation)	Scale / Range
mi05	Immigration of people fleeing war should be...	
mi06	Immigration of people persecuted for political reasons should be...	
mi07	Immigration of economic migrants should be...	1 = Fully permitted
mi08	Immigration of workers from Eastern EU countries should be...	2 = Restricted
mi09	Immigration of workers from other EU countries should be...	3 = Completely prohibited
mi10	Immigration of non-EU workers should be...	
mi11	Immigration of spouses and children of migrants should be...	
mp16inv	Refugees are a risk for the welfare state	7-Point Likert scale:
mp17inv	Refugees are a risk for public safety	1 = Strongly disagree to
mp18inv	Refugees are a risk for social cohesion	7 = Strongly agree
mp19inv	Refugees are a risk for the economy	[inversed for consistency]
pa01	In politics, where would you place yourself on a scale from left to right?	10-point Likert scale: 1 = Far left to 10 = Far right

Table 1: Overview of ALLBUS 2021 variables used in the analysis, with English item translations and response scales.

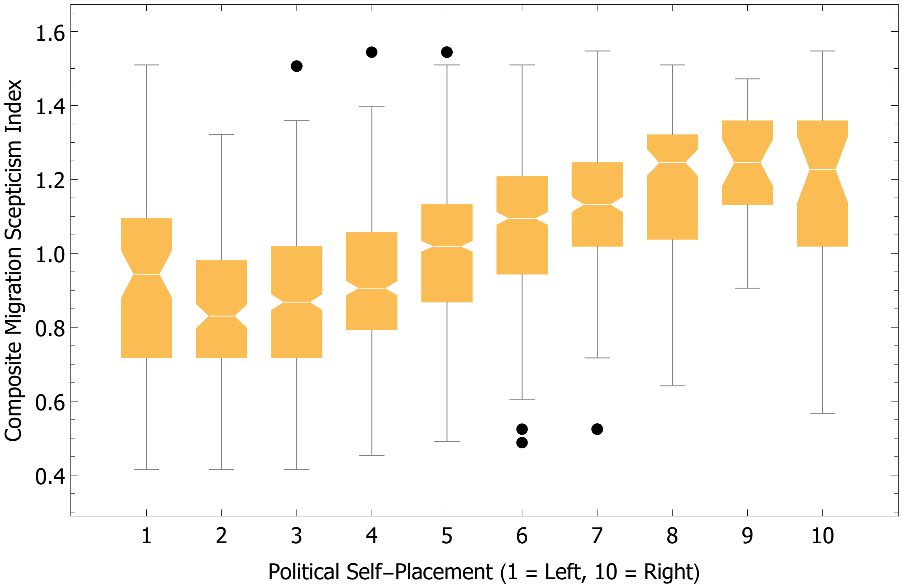


Figure 3: Distribution of migration scepticism per self-identified left-right position in the ALLBUS 2021 wave.

migration attitudes. There are 3011 respondents in the subsample relevant for the present study who answered the political self-perception question, and the questions on migration issues.

5.1.3 Perceived Polarisation

To assess the level of polarisation in the population, we calculate Ashman's D of the composite migration scepticism index. The measure is commonly used to assess the degree of separation between two overlapping distributions, and it has been proposed as a robust indicator of bimodality — a statistical property closely linked to polarisation [Tang et al., 2021], as outlined in Section 2: Bimodality captures several senses in which polarisation is typically conceptualised, including the fragmentation of a population into subgroups, the distinctness of those subgroups, and the distance between their respective positions (cf. above Section 2). Hence, we apply Ashman's D as a compact measure of the extent to which a single attitudinal dimension, such as migration scepticism, exhibits a polarised distribution. It computes the (dis)similarity of two groups as follows:

$$D = \frac{|\mu_1 - \mu_2|}{\sqrt{2 \cdot (\sigma_1^2 + \sigma_2^2)}}, \tag{3}$$

where μ_i refers to the mean score of a group i and σ_i to its standard deviation. The intuition behind this index is straightforward: Measured polarisation increases, whenever the absolute distance in the average attitude of the two groups in question increases ($|\mu_1 - \mu_2|$) or whenever the two groups get more homogeneous (i.e., the dispersion measured by σ_1 or σ_2 decreases). In our case, we derive the groups from the self-determined left-right score. This is a salient dimension of political belonging, and Windzio [2025] finds an association with attitudes towards migration in Germany, making overall political views a suitable category for reference groups. Since it is a matter of interpretation, where along this 10-point scale to split between the right-wing and left-wing cluster, we compute Ashman's D for all possible splits, and Table 2 shows the results.

Left Cluster (pa01)	**Left Size**	**Right Cluster (pa01)**	**Right Size**	**Ashman's D**
1	85	2–10	2926	0.0805
1–2	245	3–10	2766	0.1544
1–3	716	4–10	2295	0.1847
1–4	1168	5–10	1843	0.1973
1–5	1968	6–10	1043	0.1888
1–6	2512	7–10	499	0.2071
1–7	2798	8–10	213	0.2360
1–8	2946	9–10	65	0.2373
1–9	2977	10	34	0.1871

Table 2: Ashman's D benchmark values for different political self-placement splits based on `pa01`, measuring separation of migration scepticism index distributions.

5.2 Model

The descriptions for this example present a shortened overview of which specific values for attitudes, ω and ρ are used. The reader should plug them into the general framework introduced in Section 3.

5.2.1 Calibration and Network Setup

We initialise the model based on the combined migration (policy) attitude score and the left-right self-description score from the ALLBUS study 2021, as described in Section 5.1.2. We initialise each respondent with their composite migration scepticism index and political self-description on a left-right scale. This results in a population of $N = 3011$ nodes, based on the representative sample of Germany.

Despite the left-right self-description being an agent attitude, this is merely for evaluation/validation purposes. Agents perform their homophilic linkage based solely on the composite migration scepticism index. Technically, this can either be understood as two attitude dimensions $\mathbf{A} = migindex, pa01$ with $\omega_{pa01} = 0$, or as the migration scepticism index being a single attitude of interest.

The network is simulated for various homophily levels, and we use select levels to showcase behavioural patterns: $\rho = 0$ (random network) for a benchmark without endogenous polarisation, $\rho = 4$ for mild homophily, $\rho = 8$ for moderate homophily, and $\rho = 14$ for strong homophily. This parameter range is informed by a previous application in the extant literature studying inequality perceptions. Schulz et al. [2022] show that in order to capture the empirically relevant misperceptions, a ρ between 4 and 14 is necessary, with $\rho = 8$ corresponding to the average level of perceived inequality. To account for random variations in network generation, we simulate 100 Monte Carlo runs for each ρ value. Since the distribution of attitudes stems directly from the empirical input study, it is identical across all 100 Monte Carlo runs per ρ, and across all homophily levels.

5.2.2 Agents' Observations

Once the network is generated, agents make their observations. They can sense all attitudes in their own ego-network (including their own attitude). With this, they estimate their own migration attitude rank relative to their link neighbours as well as the dispersion of migration attitudes, operationalised as standard deviation in their ego-network. Furthermore, agents can sense the ego-network of one randomly selected other agent. We assume random choice of alters as a lower-bound benchmark for perceived polarisation. In empirical reality, this choice will likely not be random, as e.g. social media algorithms are typically deliberately designed to expose users to the "other side" to increase engagement Munn [2020]. Comparing this other group to their own, they calculate Ashman's D.

Note that there is no bias in agents' observations. Firstly, agents sense the true attitudes of their link-neighbours and the other randomly chosen ego-network; this could be understood to be the result of clear and unambiguous communication of attitudes. Secondly, agents correctly calculate the relevant distributional properties, i.e., they are not constrained by biases in information processing. Both these assumptions are likely violated in actual debates. However, the present study deliberately idealises here because its goal is to explore the impact of biased information availability in a polarised environment. Namely, agents take their own lived

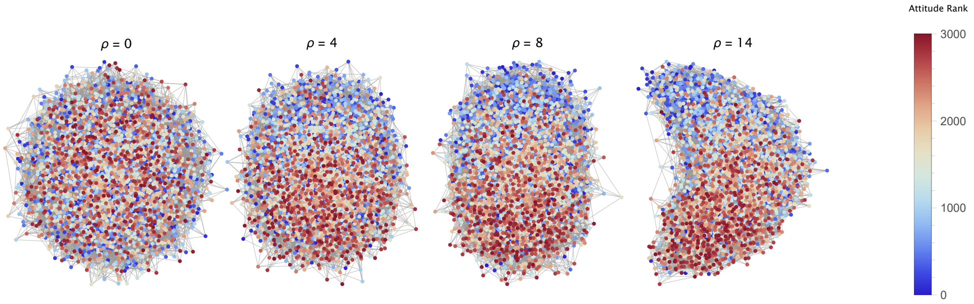


Figure 4: Exemplary graphs for different homophily levels (all taken from the first Monte Carlo Run with random seed 1). Visualisations based on Mathematica 13's "Spring Electrical Embedding" routine.

experience in their immediate social context — modelled via ego-networks — to be (representative) samples that allow inferences on the whole population.

5.3 Results

As stated above, the presentation in this section focusses on the random benchmark ($\rho = 0$), as compared to mild ($\rho = 4$), moderate ($\rho = 8$), and strong ($\rho = 14$) homophily cases. The figures in this section generally pool observations from all 100 Monte Carlo runs, such that each respondent from the survey appears 100 times, each with different network calibrations. The exemplary network graphs in Figure 4 that are based on a single random run for illustrative purposes constitute the only exception from this.

5.3.1 Polarisation of the Networks

Even mild levels of homophily cause high segregation: Figure 4 illustrates that the network for $\rho = 4$ is notably different from the random one, and that connections between agents at the lower and upper end of the attitude spectrum are rare. Yet, ties with others who have a somewhat different attitude are still fairly common. This results from the exponential form of the weighting function in Equation 1, which makes linkage to others with far distant attitudes especially unlikely, keeping the overall relative weights of those with moderately different attitudes constant. In the public forum, such connections from both extremes to moderates might create intermediaries who function as bridges between the extreme clusters. Conversely, for cases with higher homophily, agents keep more strictly to their like-minded peers, and these bridges largely disappear, making communication between extremes harder or requiring more intermediaries, as evident in the increased average shortest path length.

Such increase in average shortest path length beyond the random benchmark simultaneously means that the generated networks do not possess strong small-world properties. The networks remain connected though, and paths are only slightly longer for $\rho > 0$ than for $\rho = 0$, while clustering is much stronger, meaning that the requirements for weak small-worldiness are fulfilled [Watts and Strogatz, 1998].

5.3.2 Perception of Own Attitude Position

The first stylised fact identified in Section 5.1.1 concerns the perceived own relative attitude. To this end, they simply sort all attitudes in their ego-network and then check their own rank.

For the random network, this results in self-perceptions close to the true values, as shown in Figure 5.[6] However, even for mild homophily, there is a strong tendency towards the median, affirming the stylised fact.

5.3.3 Perceived Polarisation

The second stylised fact concerns perceptions of polarisation. We assume agents to perceive polarisation by comparing their own ego-network to the one of another randomly selected agent. This can be understood as

---

[6] The differences from the theoretical expectation of each decile containing 10% of the population are not fully matched due to the non-continuous nature of the constructed composite migration scepticism index, causing some agents with the same score to be put in different deciles.

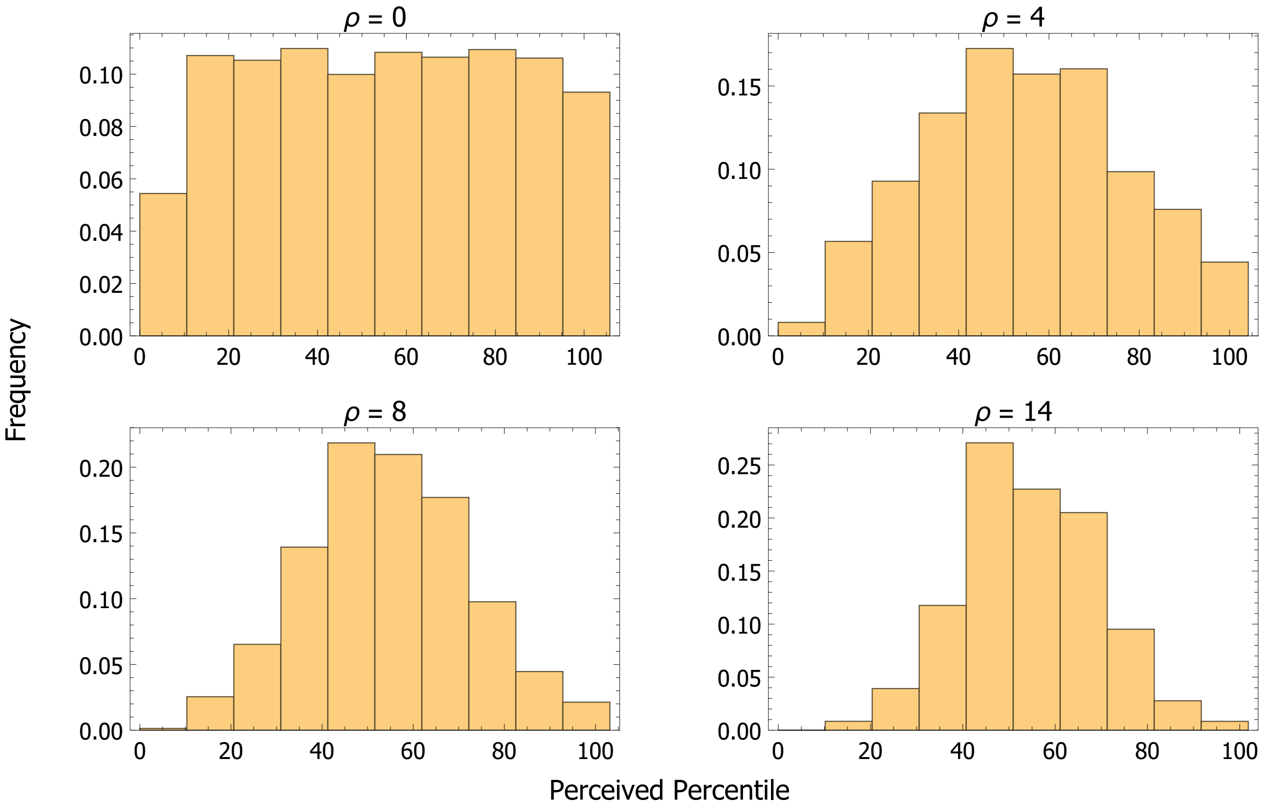


Figure 5: Perceptions of own attitude position (decile), compared to one's ego network for different homophily levels. (Trivially, accurate perception referencing the whole population would be uniformly distributed.)

people gaining insights from incidental interactions with others outside of their peer group. One might, for instance, think of such encounters as meeting strangers on the train or engaging another bubble in a (possibly controversial) social media debate. Indeed, especially for the case of social media, random choice of alters results in a lower-bound estimate for perceived polarisation, as discussed above. Based on the two groups, agents compute Ashman's D with the formula introduced in Equation 3. Hence, like for the self-perceptions above, the sensing and processing are unbiased, but the sample which agents use is biased.

Figure 6 shows the individual perceptions, comparing them to the left-right benchmark values from Table 2. While people almost underestimate polarisation in the case of a random network, homophily causes agents to perceive polarisation to be (sometimes drastically) higher than it actually is, in line with the stylised empirical fact.

The perception patterns follow a U-shape, meaning that those with attitudes further from the mean not only contribute more to the actual polarisation but also to its overestimation. The mechanism driving this shape is the combination of homophily in the generation of ego-networks with randomly selected ego-networks as a comparison group. Hence, the difference between ego and alter tends to be larger for those with extreme attitudes. For similar reasons, the dispersion of perceptions in the population as a whole and for a given attitude rises in the homophily level: For higher ρ, ego-networks grow more homogenous, making it more influential which other group is selected for comparison.

5.3.4 Homogeneity of Ego-Networks

The last stylised fact relates to the idea that polarisation inhibits openness and keeps one from taking attitudes that differ from one's own seriously. To test this fact, we draw on bounded confidence approaches and take individuals to establish confidence intervals around their own attitude: people/attitudes within this confidence interval are taken seriously by the agent in question, the ones outside it are not. To set the size of their confidence interval, an agent could refer to what the peers in their ego-network believe and assume this to be the range of normal/acceptable beliefs in society. For example, agents may take attitudes seriously that are within one standard deviation of their ego-network below or above their own.

Figure 7 displays the standard deviation of ego-networks. As expected, the ego-networks in homophilic networks show less plurality than the population as a whole. The population standard deviation corresponds to a confidence interval within which social discourse can reach consensus — for example, in the Hegselmann-Krause bounded confidence model. However, if agents instead determine their confidence interval based on their

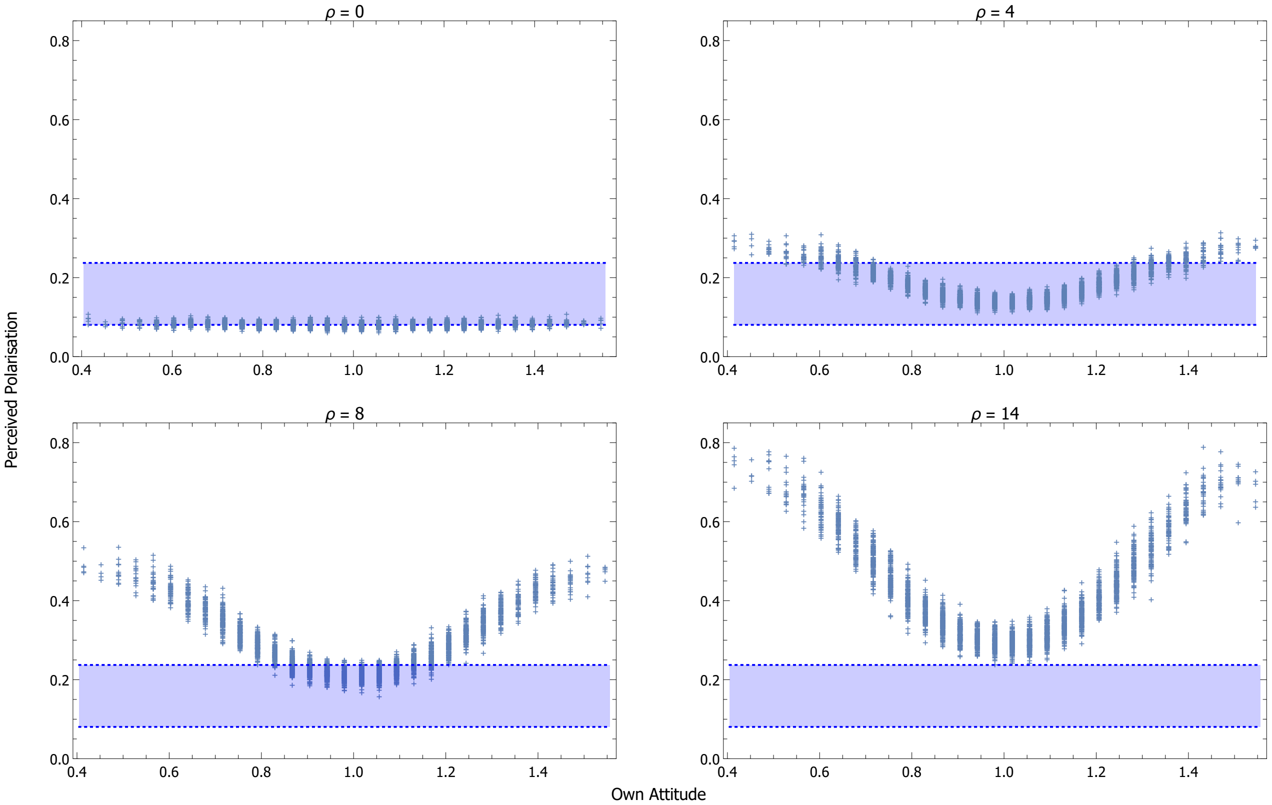


Figure 6: Perceived polarisation (Ashman's D) for different homophily levels. The marked blue area shows the range of objective polarisation, based on the left-right scale.

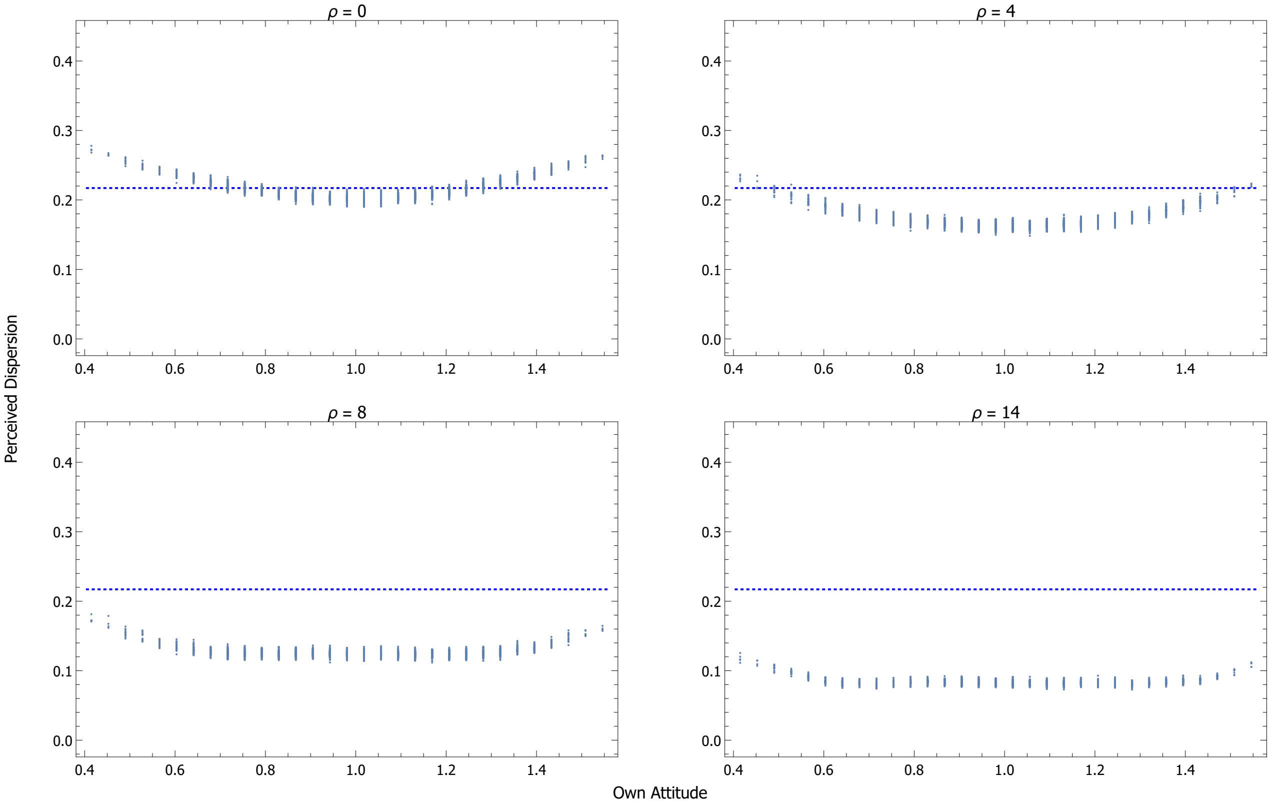


Figure 7: Perceived dispersion in attitudes compared to actual dispersion, measured by the standard deviation of attitudes. Population standard deviation as reference.

ego-network, communication dynamics would solidify polarisation [Hegselmann et al., 2002]. To get a more detailed picture, future studies could run the actual confidence dynamics interaction on the different graphs; yet even the sketch provided here showcases that homophily can inhibit fruitful debate and hence contribute to polarisation of networks as well as attitudes.

6 Discussion

This paper presented a generative algorithm for simulating network polarisation based on attitudinal homophily applicable to both undirected and directed link formation. In essence, it applies the notion of preferential attachment to node properties other than degree, aiding intuitive communication within and beyond the network science community. This proposed algorithm can work with single or multiple attitude dimensions with flexible weights; it allows for heterogeneity in the simulated population regarding these weights as well as the homophily strength, and for the specification of a minimum degree. The generated networks share key properties of Random Geometric Graphs, making their analysis with established methods straightforward. Furthermore, they commonly possess features of real-world social networks such as (weak) small-worldiness with high clustering and local cohesion as well as a right-skewed degree distribution. The algorithm's output networks can be used to evaluate structural effects of homophily, to generate how-possibly explanations of stylised empirical facts about polarisation in attitudes, and to simulate social influence processes.

The exemplary application of the algorithm to migration attitudes and polarised networks in Germany illustrates its usefulness in linking stylised facts to generative mechanisms. Simulations based on real-world attitude distributions reproduce observed patterns of polarisation and perception bias: agents in homophilic networks overestimate polarisation, misperceive their own attitudinal position, and experience reduced attitudinal diversity in their ego-networks. Hence, individual perceptions of societal attitudes may arise without biased cognition solely from the structure of social interaction. Assuming that the generative mechanism we propose is indeed (partially) operational in empirical reality, the resulting synthetic network can then also be used to identify promising targets for educational interventions with potentially significant spill-over effects.

Generally, one of the key contributions of the algorithm lies in demonstrating that complex cognitive, psychological, or ideological assumptions are neither necessary for polarised networks to emerge, nor for biased individual assumptions about attitudes and polarisation. Instead, simple homophilic preferences in link formation can generate segregated network structures which, in turn, yield biased individual samples and thus misperceptions. Therefore, the algorithm can help shift the analytical focus from individual traits to the structure in which attitudes form, circulate, and solidify. Its simplicity makes application versatile: it isolates the effect of homophily, and it offers a clean baseline for extensions, in terms of dynamic rewiring, endogenously shifting attitudes, or more nuanced sensing and cognition. Lastly, the algorithm is straightforwardly applicable to raw attributes rather than attribute differences (by dropping references to self in Equations 1 and 2) and for example capture tendencies to listen to populists with more extreme attitudes.

References

D. Abrams, M. Wetherell, S. Cochrane, M. A. Hogg, and J. C. Turner. Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology*, 29(2):97–119, 1990.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1): 47–97, Jan. 2002.

A. Alesina and M. Tabellini. The political effects of immigration: Culture or economics? *Journal of Economic Literature*, 62(1):5–46, Mar. 2024.

E. Anderson. Epistemic Bubbles and Authoritarian Politics. In E. Edenberg and M. Hannon, editors, *Political Epistemology*, pages 11–30. Oxford University Press, Oxford, 2021.

A. Asgharpourmasouleh, M. Fattahzadeh, D. Mayerhoffer, and J. Lorenz. *On the Fate of Protests: Dynamics of Social Activation and Topic Selection Online and in the Streets*, page 141–164. Springer International Publishing, Nov. 2019.

B. N. Bakker and Y. Lelkes. Putting the affect into affective polarisation. *Cognition and Emotion*, 38(4):418–436, May 2024.

R. S. Baron, S. I. Hoppe, C. F. Kao, B. Brunsman, B. Linneweh, and D. Rogers. Social Corroboration and Opinion Extremity. *Journal of Experimental Social Psychology*, 32(6):537–560, Nov. 1996.

E. Begby. From Belief Polarization to Echo Chambers: A Rationalizing Account. *Episteme*, pages 1–21, 2022.

M. Billig and H. Tajfel. Social categorization and similarity in intergroup behaviour. *European Journal of Social Psychology*, 3(1):27–52, 1973.

A. Bonica, A. S. Chilton, and M. Sen. The Political Ideologies of American Lawyers. *Journal of Legal Analysis*, 8(2):277–335, Dec. 2016.

C. Bothorel, J. D. Cruz, M. Magnani, and B. Micenková. Clustering attributed graphs: Models, measures and methods. *Network Science*, 3(3):408–444, Mar. 2015.

A. Bruns. *Are Filter Bubbles Real?* Polity, Cambridge, Aug. 2019.

E. Burnstein and A. Vinokur. Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology*, 13(4):315–332, July 1977.

J. Dall and M. Christensen. Random geometric graphs. *Physical Review E*, 66(1), July 2002.

A. Dasgupta, R. Kumar, and D. Sivakumar. Social sampling. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, page 235–243, New York, NY, USA, 2012. Association for Computing Machinery.

G. Deffuant, F. Amblard, G. Weisbuch, and T. Faure. How can extremism prevail? a study based on the relative agreement interaction model. *Journal of Artificial Societies and Social Simulation*, 5(4), 2002.

M. H. Degroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, Mar. 1974.

A. S. Dippel, L. Hetzer, and A. M. Burger. Links oder rechts? die ideologische selbstverortung von wähler:innen und ihre wahrnehmung von parteien in deutschland. *easy social sciences*, 2022.

A. M. Enders and M. T. Armaly. The differential effects of actual and perceived polarization. *Political Behavior*, 41:815–839, 2019.

J. M. Epstein. Agent-based computational models and generative social science. *Complexity*, 4(5):41–60, 1999.

M. Esteve-Del-Valle. Homophily and polarization in twitter political networks: A cross-country analysis. *Media and Communication*, 10(2), Apr. 2022.

E. J. Finkel, C. A. Bail, M. Cikara, P. H. Ditto, S. Iyengar, S. Klar, L. Mason, M. C. McGrath, B. Nyhan, D. G. Rand, L. J. Skitka, J. A. Tucker, R. Saxe, J. J. Van Bavel, C. S. Wang, and J. N. Druckman. Political sectarianism in America. *Science*, 370(6516):533–536, 2020.

M. Galesic, H. Olsson, and J. Rieskamp. Social sampling explains apparent biases in judgments of social environments. *Psychological Science*, 23(12):1515–1523, Oct. 2012.

GESIS-Leibniz-Institut Für Sozialwissenschaften. Allgemeine bevölkerungsumfrage der sozialwissenschaften allbus - kumulation 1980-2021, 2024.

C. Hare. Constrained Citizens? Ideological Structure and Conflict Extension in the US Electorate, 1980–2016. *British Journal of Political Science*, 52(4):1602–1621, Oct. 2022.

R. Hegselmann, U. Krause, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3):1–33, 2002.

A. D. Henry, D. Mitsche, and P. Prałat. Homophily, influence and the decay of segregation in self-organizing networks. *Network Science*, 4(1):81–116, Mar. 2016.

D. Hirschman. Stylized facts in the social sciences. *Sociological Science*, 3:604–626, 2016.

S. B. Hobolt, K. Lawall, and J. Tilley. The Polarizing Effect of Partisan Echo Chambers. *American Political Science Review*, 118(3):1464–1479, Aug. 2024.

M. J. Hornsey. Social Identity Theory and Self-categorization Theory: A Historical Review. *Social and Personality Psychology Compass*, 2(1):204–222, 2008.

G. A. Huber and N. Malhotra. Political Homophily in Social Relationships: Evidence from Online Dating Behavior. *The Journal of Politics*, 79(1):269–283, 2017.

L. Huddy, L. Mason, and L. Aarøe. Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity. *American Political Science Review*, 109(1):1–17, Feb. 2015.

S. Iyengar. E Pluribus Pluribus , or Divided We Stand. *Public Opinion Quarterly*, 80(S1):219–224, Jan. 2016.

S. Iyengar and S. J. Westwood. Fear and Loathing across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 59(3):690–707, 2015.

S. Iyengar, T. Konitzer, and K. Tedin. The Home as a Political Fortress: Family Agreement in an Era of Polarization. *The Journal of Politics*, 80(4):1326–1338, Oct. 2018.

S. Iyengar, Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood. The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, 22(Volume 22, 2019):129–146, May 2019.

N. Kaldor. *Capital Accumulation and Economic Growth*, page 177–222. Palgrave Macmillan UK, 1961.

D. Klein and J. Marx. Wenn du gehst, geh ich auch. die rolle von informationskaskaden bei der entstehung von massenbewegungen. *Politische Vierteljahresschrift*, 58(4):560–592, 2017.

T. Kuran. Sparks and prairie fires: A theory of unanticipated political revolution. *Public choice*, 61(1):41–74, 1989.

E.-J. Lee. Deindividuation Effects on Group Polarization in Computer-Mediated Communication: The Role of Group Identification, Public-Self-Awareness, and Perceived Argument Quality. *Journal of Communication*, 57 (2):385–403, June 2007.

Y. Lelkes. Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly*, 80(1):392–410, Jan. 2016.

P. MacCarron, K. Kaski, and R. Dunbar. Calling Dunbar's numbers. *Social Networks*, 47:151 – 155, 2016.

C. F. Manski and D. McFadden. *Structural analysis of discrete data with econometric applications*. MIT press Cambridge, MA, 1981.

L. Mason. *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press, Chicago, Apr. 2018.

L. Mason and J. Wronski. One Tribe to Bind Them All: How Our Social Group Attachments Strengthen Partisanship. *Political Psychology*, 39(S1):257–277, 2018.

D. M. Mayerhoffer and J. Schulz. Perception and privilege. *Applied Network Science*, 7(1), 2022.

M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, Aug. 2001.

I. D. Mehlhaff. A Group-Based Approach to Measuring Polarization. *American Political Science Review*, 118(3): 1518–1526, Aug. 2024.

B. V. Meylahn and C. Searle. Opinion dynamics beyond social influence. *Network Science*, 12(4):339–365, 2024. doi: 10.1017/nws.2024.14.

S. Moscovici and M. Zavalloni. The group as a polarizer of attitudes. *Journal of Personality and Social Psychology*, 12(2):125–135, 1969.

L. Munn. Angry by design: toxic communication and technical architectures. *Humanities and Social Sciences Communications*, 7(1):1–11, 2020.

D. G. Myers and G. D. Bishop. Discussion Effects on Racial Attitudes. *Science*, 169(3947):778–779, Aug. 1970.

D. G. Myers, J. B. Bruggink, R. C. Kersting, and B. A. Schlosser. Does Learning Others' Opinions Change One's Opinions? *Personality and Social Psychology Bulletin*, 6(2):253–260, June 1980.

A. Nasuto and F. Rowe. Understanding anti-immigration sentiment spreading on twitter. *PLOS ONE*, 19(9): e0307917, Sept. 2024.

M. E. J. Newman and R. M. Ziff. Fast monte carlo algorithm for site or bond percolation. *Physical Review E*, 64 (1), June 2001.

B. Nyhan, J. Settle, E. Thorson, M. Wojcieszak, P. Barberá, A. Y. Chen, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Dimmery, D. Freelon, M. Gentzkow, S. González-Bailón, A. M. Guess, E. Kennedy, Y. M. Kim, D. Lazer, N. Malhotra, D. Moehler, J. Pan, D. R. Thomas, R. Tromble, C. V. Rivera, A. Wilkins, B. Xiong, C. K. de Jonge, A. Franco, W. Mason, N. J. Stroud, and J. A. Tucker. Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972):137–144, Aug. 2023.

E. Pariser. *The Filter Bubble: What The Internet Is Hiding From You.* Penguin, London, May 2011.

A. Reutlinger, D. Hangleiter, and S. Hartmann. Understanding (with) toy models. *The British Journal for the Philosophy of Science*, 69(4):1069–1099, dec 2018.

S. Schneider-Strawczynski and J. Valette. Media coverage of immigration and the polarization of attitudes. *American Economic Journal: Applied Economics*, 17(1):337–368, Jan. 2025.

J. Schulz and D. M. Mayerhoffer. A network approach to expenditure cascades. *Review of Behavioral Economics*, 10(3):229–262, 2023.

J. Schulz, D. M. Mayerhoffer, and A. Gebhard. A network-based explanation of inequality perceptions. *Social Networks*, 70:306–324, 2022.

M. A. Schulz and S. Scheller. Being a believer: Social identity in post-truth political discourse. *Inquiry: An Interdisciplinary Journal of Philosophy*, 2024.

M. Schwaninger and S. Grösser. System dynamics as model-based theory building. *Systems Research and Behavioral Science*, 25(4):447–465, July 2008.

J. Sides, M. Tesler, and L. Vavreck. *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America.* Princeton University Press, Princeton, Oct. 2018.

F. Squazzoni. *Agent-based computational sociology.* John Wiley & Sons, 2012.

D. Steppat, L. Castro Herrero, and F. Esser. Selective exposure in different political information environments – How media fragmentation and polarization shape congruent news use. *European Journal of Communication*, 37(1):82–102, Feb. 2022.

R. Sugden. Credible worlds, capacities and mechanisms. *Erkenntnis*, 70(1):3–27, 2009.

C. R. Sunstein. The Law of Group Polarization. *Journal of Political Philosophy*, 10(2):175–195, 2002.

C. R. Sunstein. *Going to Extremes: How Like Minds Unite and Divide.* Oxford University Press, Oxford, May 2009.

C. R. Sunstein. Republic: Divided Democracy in the Age of Social Media. pages 1–328, 2018.

F. Taghikhah, T. Filatova, and A. Voinov. Where does theory have it right? a comparison of theory-driven and empirical agent based models. *Journal of Artificial Societies and Social Simulation*, 24(2), 2021.

H. Tajfel and J. C. Turner. The Social Identity Theory of Intergroup Behavior. In S. Worchel and W. G. Austin, editors, *Psychology of intergroup relations*, pages 7–24. Nelson, Chicago, 1986.

R. B. Talisse. *Overdoing Democracy: Why We Must Put Politics in its Place.* Oxford University Press, New York, 2020.

R. B. Talisse. Problems of Polarization. In E. Edenberg and M. Hannon, editors, *Political Epistemology*, pages 209–225. Oxford University Press, Oxford, 2021.

T. Tang, A. Ghorbani, F. Squazzoni, and C. G. Chorus. Together alone: a group-based polarization measurement. *Quality amp; Quantity*, 56(5):3587–3619, Dec. 2021.

C. Turner. Online echo chambers, online epistemic bubbles, and open-mindedness. *Episteme*, page 1–26, Nov. 2023.

P. Ueffing, F. Rowe, and C. H. Mulder. Differences in attitudes towards immigration between australia and germany: The role of immigration policy. *Comparative Population Studies*, 40(4), Dec. 2015.

S. Vaidhyanathan. *Antisocial Media: How Facebook Disconnects Us and Undermines Democracy.* Oxford University Press, 2018.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

M. Windzio. Limits of rationality? moral foundations and polarization in attitudes towards diversity in germany. *Rationality and Society*, Mar. 2025.

D. Wollebæk, R. Karlsen, K. Steen-Johnsen, and B. Enjolras. Anger, fear, and echo chambers: The emotional basis for online behavior. *Social Media + Society*, 5(2), Apr. 2019.